Trends in Plant Science

CellPress REVIEWS

Feature Review

Challenges and Opportunities in Machine-Augmented Plant Stress Phenotyping

Arti Singh,[1,*] Sarah Jones,[1] Baskar Ganapathysubramanian,[2] Soumik Sarkar,[2] Daren Mueller,[3] Kulbir Sandhu,[1] and Koushik Nagasubramanian[4]

Plant stress phenotyping is essential to select stress-resistant varieties and develop better stress-management strategies. Standardization of visual assessments and deployment of imaging techniques have improved the accuracy and reliability of stress assessment in comparison with unaided visual measurement. The growing capabilities of machine learning (ML) methods in conjunction with image-based phenotyping can extract new insights from curated, annotated, and high-dimensional datasets across varied crops and stresses. We propose an overarching strategy for utilizing ML techniques that methodically enables the application of plant stress phenotyping at multiple scales across different types of stresses, program goals, and environments.

Highlights

Plant stress phenotyping is challenging to implement at multiple organizational scales (leaf, canopy, field).

There is a need to improve the speed, accuracy, reliability, and scalability of stress phenotyping while allowing flexibility for highly variable program goals.

Advances in ML algorithms create opportunities for augmented plant stress phenotyping to address these challenges.

Understanding Plant Stress Is Crucial for Yield Protection

Plant stress is a state of plant growth under non-ideal environmental conditions caused by various biotic (pathogen, insect, pest, and weed) and abiotic (temperature stress, nutrient deficiency, toxicity, herbicide) factors. Significant crop yield loss due to various plant stresses has the potential to threaten global food security [1]. Plant disease epidemics are a constant threat and continue to emerge owing to complex host–pathogen–environment dynamics [2,3]. Global climate change can exacerbate this situation because of the predicted increases in insect and pathogen pressure for major grains including rice (*Oryza sativa*), maize (*Zea mays* L.), and wheat (*Triticum aestivum*) [4]. Moreover, weather-related challenges such as drought, flooding, hail, and windstorms adversely affect crop production. Yield preservation and protection is a dynamic challenge for pathologists, entomologists, plant breeders, and crop producers globally. Understanding plant stress is crucial for improving yield protection to meet with the growing demand for food production [5]. In the past decade, significant advances in image processing and **machine learning** (ML; see Glossary) algorithms have been made to handle image-based stress datasets for automated data analysis and application of trained models [6–8]. We review the development and application of ML algorithms for image-based plant stress phenotyping at multiple scales ranging from leaf and canopy (plot) to field (production) scale. We discuss some of the major challenges in the practical application of ML algorithms, and list future efforts that will be necessary to make ML a more mainstream tool in plant stress phenotyping applications and usage.

Foundations in Measuring Plant Stress

Plant phenotyping refers to the set of methodologies and protocols that are used to measure plant traits under distinct environmental conditions with a specified **accuracy** and **precision** at different scales of organization, from organs to canopies [9,10]. Precision agriculture refers to a farm management strategy of monitoring and responding to spatial heterogeneities within crop stands [10]. Foliar plant stress phenotyping is integral to plant phenotyping and precision in agriculture, and occurs at a variety of scales from the single-leaf scale to higher scales of plant canopies and whole fields. We use the term ‘plant stress severity’ to encompass both biotic and abiotic stresses. Plant stress severity measurement is important for evaluating management

[1]Department of Agronomy, Iowa State University, Ames, IA, USA
[2]Department of Mechanical Engineering, Iowa State University, Ames, IA, USA
[3]Department of Plant Pathology and Microbiology, Iowa State University, Ames, IA, USA
[4]Department of Electrical and Computer Engineering, Iowa State University, Ames, IA, USA

*Correspondence: arti@iastate.edu (A. Singh).

approaches, plant breeding selection strategies, and testing new varieties for their ability to mitigate crop losses. Therefore, advances in the accuracy and automation of stress severity ratings can improve the rate of genetic gain across crops and lead to the development of new management strategies [11,12].

Plant stress assessments quantify the visible signs or symptoms of stress and its progression on individual plant units (e.g., leaf, stem, or roots) at the leaf, canopy, plot, and field levels. Disease quantification has traditionally been captured through visual assessment of incidence, prevalence, and severity. **Disease incidence** is captured in two ways: (i) single-plant scale: the proportion of diseased plant units among the total number of plant units on a single plant, or (ii) multiple-plant scale: the number of diseased plants among the total number of plants in a plot or field [5]. **Disease prevalence** is the percentage of a larger geographic area where the disease has been detected [13,14]. **Disease severity** refers to the area of plant tissue affected by the disease (commonly presented as a percentage) on a leaf or on the entire plant canopy. The combination of these measurements reveals how many plants are infected, the severity of the stress, and the geographic spread of the stress. Such definitions may also be applied to insect damage or abiotic stresses such as herbicide damage. The decision to measure disease incidence, prevalence, or severity depends on the stress being studied, its expected intensity, and the objectives of the study [15]. The quality of stress assessments is judged on two main metrics: accuracy and precision. Accuracy refers to the extent that the estimated stress assessment corresponds to the definite value of stress [14,16]. Precision measures the amount of variability, or closeness, among repeated measurements, and is often called **reliability** in plant disease assessments [15,17]. Precision, therefore, characterizes the repeatability/reproducibility of the assessment.

The evaluation of a stress measurement system for both qualitatively and quantitatively expressed traits requires utilization of these metrics, which also allows standardization across scales. Unlike many qualitative traits that are obvious and understandable in their expression, quantitative stress-resistance traits add challenges to the phenotyping process. Quantitative traits rarely express in discrete infection categories, and often develop in a continuous distribution, complicating stress quantification. Several rating scales, including nominal (descriptive), ordinal, interval (categorical) and ratio scales, have been developed to handle the complexities of rating quantitative traits [5]. These different scales describe continuous or discrete characteristics, and may be general or specific to a single host–pathogen relationship. This is where leveraging ML approaches could result in more robust methodologies, especially if severity-based scales are used in model development. Scales may also be supplemented with **standard area diagrams** (SADs), which will be discussed in more depth later.

Field scouting with visual stress phenotyping is the most common method in stress observation owing to convenience and cost [5,15]. However, significant drawbacks are associated with visual estimation methods [5,18,19]. Measures of agreement, **inter-rater reliability**, and **intra-rater reliability** all indicate that there are differences among individual raters as well as variability among assessments by a single rater [20]. In general, estimates made by experienced raters are more accurate and reliable compared with those made by inexperienced raters [21–23]. Some individuals can be inherently biased, which causes some disease severity values to be chosen over others compared with random estimations around the actual disease severity. Bias is particularly acute at low levels of disease (<10%). In addition, raters tend to overestimate the disease severity level [5,22,24,25]. For many applications, such as screening for resistance in breeding nurseries, accurate assessments at low severity levels are particularly important. The difficulty of determining the level of infection combined with human error compounds the problem of inaccurate ratings.

Glossary

Accuracy: the degree of the closeness of an estimate (disease assessment) to the true value or standard [14].
Active learning (AL): a smart annotation system that utilizes ML to select the most informative data samples for an expert to annotate so as to streamline annotation and improve quality of training datasets.
Artificial intelligence (AI): a branch of computer science based on the idea that human learning can be defined and mimicked by computers. The computer systems developed are capable of emulating human intelligence in learning and problem solving.
Deep learning (DL): a class of ML algorithms that receives its name from the deep set of layers, or a stack of multiple processing layers, where output of one layer is used as the input in the following layer, enabling higher-level details to be discerned from complex datasets.
Disease incidence: the rate of occurrence, or percentage, of diseased plant units (whole plant, leaves, roots, stems) among all plant units sampled for the presence of a disease.
Disease prevalence: a larger-scale measurement of disease incidence that characterizes the percentage of a population affected by a disease at the field, county, or even state level of a specific geographic area.
Disease severity: the relative or absolute area of a plant sampling unit (leaf, stem, fruit, etc.) that exhibits disease symptoms.
Inter-rater reliability (reproducibility): the statistical similarity of disease measurements taken from the same sampling unit by different raters.
Intra-rater reliability (repeatability): the statistical similarity of repeated measurements taken from the same sampling unit by the same rater (or device) under different conditions.
Machine learning (ML): a class of methods used in AI where an algorithm can automatically adjust itself, or learn, based on incoming information or experience.
Precision: the amount of variability among measurements, or the degree of reliability and repeatability of a measurement; note that highly precise measurements are not necessarily accurate.
Reliability: the extent to which the same measurement obtained under

High-Throughput Plant Stress Phenotyping

Improvements in automated phenotyping systems that are integrated with **artificial intelligence** (AI) could be a panacea for problems associated with the current state-of-the-art plant stress phenotyping. The advent of sophisticated sensors, high-throughput phenotyping (HTP), and advanced data analytics through ML, as well as the ubiquitous availability of computational infrastructure and resources, now provide a promising path forward to resolve challenges with plant stress phenotyping [26]. Challenges include the need to remove the remaining subjectivity, deepen the quality of data collection, and improve scalability and versatility. In this regard, high-throughput AI technologies for early stress detection before visual symptoms [27] may be the next step for precision agriculture, field-based application, and eventually research in small test plots.

Big data collection enabled by unmanned aerial system (UAS) technology and ground robots coupled with ML will add value to agricultural technologies based on field stress phenotyping to improve farmers' decision-making power and crop yields. However, there is a need to improve and standardize plant stress data-collection protocols. Furthermore, a seamless pipeline of data acquisition, curation, and analytics should be developed to deploy ML-based real-time plant stress phenotyping on a large scale. The advantage of using ML for data analytics is that, once a robust ML framework (with high accuracy and precision) is trained on a large image dataset spanning all the variability present for that particular trait, it can be easily packaged into an intuitive graphical user interface deployable on a smartphone, robot, and UAS for routine use in field applications (Figure 1). The current dependency of advanced ML methods on a large, labeled dataset for the development of accurate ML models creates a significant challenge for the plant science community. This is because labeling a diverse dataset requires a significant resource investment by trained plant scientists, and is usually a difficult proposition. However, recent advances in ML methods (e.g., **active learning**, AL [28]; **semi-supervised learning** [29]; and self-supervised learning [30]) open up the possibility of settling with smaller labeled datasets. The diffusion of these novel developments by the ML community to the plant science community requires close collaboration between the engineering and agriculture disciplines.

different conditions yields similar results [133].
Semi-supervised learning: assists in situations where the labeled training data are too few to design a high-quality algorithm by using a small amount of labeled data and a large amount of unlabeled data in training.
Standard area diagrams (SADs): also known as diagrammatic scales, disease scales, and disease diagrams. SADs are visual aids in pictorial or graphical form for disease severity ratings showing examples of the classes to be categorized from the scale of disease severity levels displayed by crop disease symptomology.
Supervised learning: training data are labeled with a known identity when training a ML algorithm.
Transfer learning: a technique used in ML to optimize learning efficiency by repurposing a model that was developed for one objective as the starting point for a second objective through the transfer of learned knowledge or features from the first model to the second. This also assists in situations where the second model has fewer training data for successful training from scratch.
Unsupervised learning: commonly uses clustering to divide pixels into groups that are not previously specified without the assistance of labeled data.

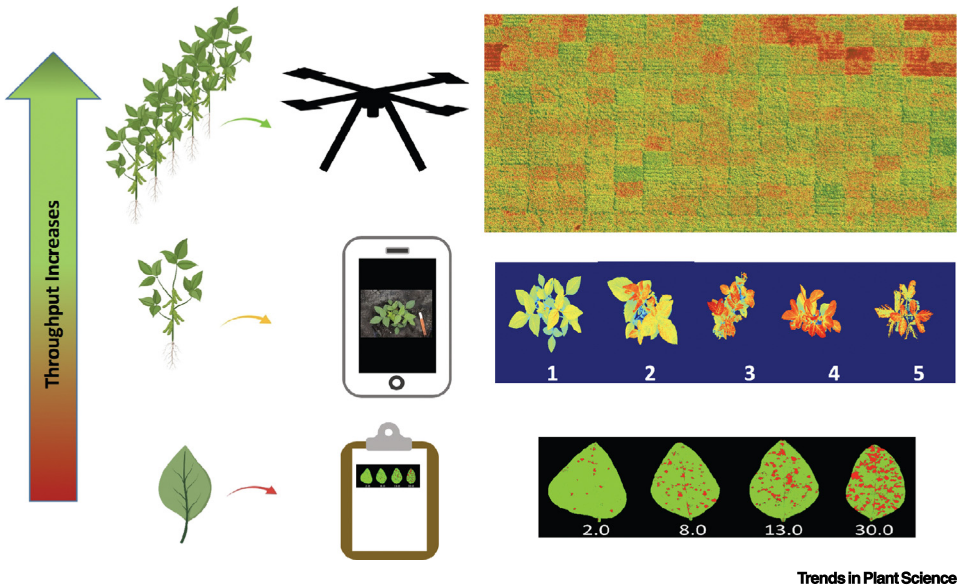

Figure 1. The Hierarchical Scale of Plant Stress Phenotyping. It can be managed through implementation of new technology to develop an overarching streamlined and higher-throughput system of plant stress phenotyping.

Automated Plant Stress Phenotyping by ML

ML and its subtype **deep learning** (DL) expand our ability to extract information collected by HTP systems [26,31]. With the advent of big data supported by higher-throughput phenotyping systems, ML approaches are applicable to multiple areas of plant stress phenotyping including identification, classification, quantification, and prediction (ICQP) [26]. Recent studies address the identification and classification aspects of ICQP paradigm for plant stress phenotyping using ML tools; however, few studies have utilized ML for quantification and prediction of plant disease.

Supervised learning comprises a class of algorithms that analyze prelabeled training datasets to generate a function that is used for mapping previously unseen data into already established labels. Supervised classifiers/regressors use prelabeled training data to distinguish patterns in the data that are associated with the labels [32]. The training examples in supervised learning methods are labeled with their known and expected outputs [26,33]. A semi-supervised learning approach can also be used for training the ML algorithm by only labeling a partial set of the training data [34,35]. Semi-supervised learning saves human hours (for labeling tasks) and material resources over supervised learning methods. Both supervised and semi-supervised classical ML (non-DL) approaches require manual feature extraction, in other words the selection of important traits from raw data to train ML algorithms. Thus, supervised and semi-supervised classical ML approaches do not preclude the involvement of human expertise and therefore can be still error-prone. Moreover, versatility may remain a challenge if the features are species-dependent and are not analogous across crop species. On the other hand, if the training data are labeled by an expert, this method provides a balance between human and machine intelligence and is very useful in numerous situations. Semi-supervised (or weakly supervised) and AL-based ML approaches still offer opportunities to utilize domain expertise to label non-randomly selected data in training a DL algorithm [36]. By contrast, **unsupervised learning** approaches do not require image annotation for training [26]. The absence of human annotation increases objectivity by removing the potential for human error as well as the costly labor involved in image annotation. In the absence of time and cost constraints of labeling training data, larger datasets can be utilized more efficiently in model training, thus increasing scalability and algorithm performance [37,38].

Automated Plant Stress Phenotyping by DL

DL is a subtype of ML in which a computer learns to extract hierarchical features and make decisions with image, text, or other forms of data. DL models typically represented as multilayer neural networks are trained to simultaneously extract features and make decisions using advanced backpropagation algorithms. Although lower layers (closer to input data) capture simpler features, upper layers (closer to decision layers) learn to capture more complex features that are composed of the simpler features [39]. The ability of DL algorithms to handle large numbers of image features and deal with high-dimensional, complex datasets, especially when using input from hyperspectral sensors, makes it particularly suitable for achieving fine-grained stress severity assessment [40,41]. Advances in DL-based super-resolution are enabling accurate (re)creation of high-resolution data at the canopy scale from high-throughput, coarse-resolution sensor data (e.g., hyperspectral data captured from UAS). This can potentially resolve the issue of resolution–throughput tradeoff by providing high-resolution data across large geographical spreads. A persistent challenge with far field-based high-throughput imaging (UAS, satellite, etc.) is that lower levels of the canopy are difficult to measure. There are several promising approaches to circumvent these, including (i) tandem measurements via synchronous deployment of ground and aerial payloads [42], and (ii) exploring the possibility of non-line-of-sight sensing [43].

Multiscale Stress Phenotyping by Machine-Based Automation

Leaf-Scale Stress Phenotyping

In the context of ML-based phenotyping, using leaves as an example, disease incidence requires 'segmenting' out individual leaves, classifying the stress state of each segmented leaf, and then computing incidence percent. This is an inherently tough problem because of difficulty in segmenting individual plant units owing to occlusion/overlap etc. On the other hand, disease severity only requires pixel-by-pixel characterization of the stress-state (without the need to segment individual plant units), followed by computing the percentage of pixels expressing a particular stress-state. This is a substantially easier proposition.

Stress phenotyping can be performed at multiple crop levels. Primary goals at the leaf level includes disease identification, classification (healthy vs stressed), and quantification of symptoms. Assistive technologies include SADs, image analysis software, and smartphone apps. SADs, also known as diagrammatic keys or diagrammatic scales, were developed to calibrate the human eye to assess the severity of the particular disease symptomatology under evaluation [44–47]. The use of SADs aims to improve accuracy and reliability, as well as to overcome subjectivity associated with traditional, unaided visual assessment techniques. Generic guidelines for development of SADs were published in the late 1960s [46,48]. To determine severity ratings, diseased tissue samples are compared with the diagram of various severity levels [49,50]. A robust SAD should be applicable even under a wide range of environmental conditions and encompass the range of severity values that would be encountered in the field [51]. Before a SAD can be recommended as a tool for severity ratings, it must be quantitatively evaluated for accuracy and precision, and any bias the SAD may cause in ratings [45]. Over the past century these pictorial diagrams, each depicting a range of distinct disease severities on individual plant parts, have been used in phytopathological studies as reference guides focused on individual plant parts such as leaves or leaflets, heads of wheat, or fruit [52–54]. Recently, Del Ponte *et al.* provided an extensive review of 127 SADs originating from 105 studies published between 1991 to 2016 covering major economic crops [55]. The positive impact of SADs on accuracy and reliability of disease estimates was reviewed by Bock *et al.* [5]. The standardization created by SADs has been shown to increase the accuracy, speed, precision, agreement, and reliability of raters, especially that of inexperienced raters [23,49,50,56–58].

More recently, plant stress severity quantification software has been developed. These include commercial programs, including 'Assess' [59] and 'QUANT' [60], and free tools such as 'ImageJ' [61], to assist plant pathologists with more accurate measurement of quantitative disease resistance through image analysis. Another automated method was developed in wheat for automated batch processing using macro in ImageJ to analyze percent lesion coverage, pycnidia size, and pycnidia density of *Zymoseptoria tritici* for more accurate and precise measurement of disease compared with visual estimates of virulence [62]. Various digital approaches have also laid the foundation for plant stress quantification software, including several smartphone applications. Leaf Doctor is an iOS-compatible mobile application that was developed to calculate disease severity on individual leaves [63]. 'Estimate' can calculate disease severity to aid real-time treatment decisions and for data collection in the field for yam anthracnose, maize streak virus, and *Cercospora* leaf spot of beet [64]. Diseased leaves are compared with the SAD in the app, and leaves that most closely resemble the field image are recorded by the app. The guidelines for SADs development, design, and testing are presented in Del Ponte *et al.* [55] (logarithmic, ordinal, continuous, etc.). These are expected to result in the highest accuracy of severity assessment [55,64]. When using smartphone applications that still rely on the human eye, training on the rating protocol and disease symptoms are required for accurate and reliable ratings [65].

To an extent, the previously discussed methods remove subjectivity in stress quantification at the individual plant unit scale (leaf, stem, root, and fruit). An improved multimodal framework can streamline HTP pipelines owing to the multistep process that achieves identification, classification, and quantification of various foliar stresses. Ghosal *et al.* approached a twofold problem in soybean by utilizing a deep convolutional neural network (DCNN) model and explanation framework to first identify the foliar stress appearing in each leaf image, and to then reveal what symptoms were used by the model for classification to severity quantification [6]. The model accurately identified different soybean [*Glycine max* L. (Merr.)] stresses including bacterial and fungal diseases, nutrient deficiencies, and herbicide injury with an overall accuracy of 94% on the testing dataset. To uncover the features learned by the algorithm for classification, a top-K high-resolution feature map was produced to discern which features were chosen by the DCNN for classification. The explanation map-based framework, xPLNeT, increases confidence in the model by looking under the hood of a 'black box system' and allows unsupervised severity quantification. Such explanation techniques may relieve the confidence barrier that scientists may have towards using 'black box' models. However, these models were built using >60 000 images and even with data augmentation techniques, and it is a non-trivial task to assemble such large plant stress datasets.

The necessity for large volumes of training data for stress identification and quantification by focused ML algorithms makes collecting and curating datasets crucial [66]. PlantVillage is a repository developed by plant scientists who contributed images on plant health (containing 87 848 photos of both healthy and diseased leaves of 25 species) and enables the development of mobile disease diagnostics [67]. DCNN and other ML models have been exhaustively applied to the PlantVillage dataset to identify tomato diseases [68–71] and to identify diseases in multiple crop species simultaneously [67,72,73]. Although the PlantVillage plant stress dataset is large, there are still limitations, such as limited diversity in severity, imaging platforms, and sampling variation of the image set in each disease group. Independently of PlantVillage datasets, Boulent *et al.* reviewed 19 studies integrating DCNN approaches for automatic plant disease identification from leaf images [74]. Other successful applications of convolutional neural network (CNN) methods include classification of diseased apple leaves into severity level categories (healthy stage, early stage, middle stage, and end stage) [75] and segmentation of powdery mildew symptoms on cucumber leaf images for severity quantification [76]. Although the routine use of ML-based plant stress phenotyping is still limited to research settings, there are several examples of ML application in production fields. One such application, 'Nuru', is an AI-based app embedded in the PlantVillage app to detect cassava disease, fall army worm, and other stresses, and is reported to be currently used by African farmers (https://plantvillage.psu.edu/solutions#nuruF). Another example is citrus greening disease Huanglongbing (HLB) caused by the bacterial pathogen *Candidatus Liberibacter asiaticus*. The optical signatures captured by miniature spectrophotometers are analyzed by a cloud-based AI algorithm that diagnoses HLB months before it is visible to the human eye, allowing the infected tree to be expeditiously uprooted to prevent transmission of infection to neighboring trees (www.croptix.solutions/).

Plant Canopy and Small-Plot Stress Phenotyping

The primary goal for plant canopy/small-plot stress phenotyping is to collect stress response data in situations where single-leaf phenotypes alone would not provide sufficient information. Such traits include canopy size, height, canopy structure, and branching, or traits that may vary in severity within a single plant canopy because these are stress manifestations that affect yield. Traits such as drought wilt score in soybean are estimated visually, often according to differing scales, thus adding complexity to the analysis and complicating collaboration between programs [77,78]. Large-scale training data that reflect the complexity of the target environment are required, especially to develop more robust models that circumvent background heterogeneity, intra-class variability, image acquisition background and lighting conditions, leaf angle variation, and multi-symptom development in the field [74]. CLS Rater, a computer vision system involving supervised training, rated disease severity with refined accuracy compared with human raters on a numerical scale from RGB (red/green/blue) images collected by a tractor-mounted camera in small-plot canopies under field conditions [79]. An automated phenotyping workflow integrating imaging, data analytics, and ML was developed to determine severity level of iron-deficiency chlorosis (IDC), which causes interveinal chlorosis and necrosis in soybean [7]. Within the work-flow, (i) completely automated image preprocessing was used for segmenting the plant canopy from the background, (ii) feature extraction determined percent yellow and percent brown pixels in the plant canopy image, and (iii) a hierarchical classification-based supervised ML model using linear discriminant analysis and support vector machines (SVMs) was used to classify each canopy image on an ML-derived 1–5 rating score and ML-derived severity on a scale of 1–100%. This automated workflow was designed for implementation as a smartphone application. The SVM-based hierarchical classifier was evaluated in a genome-wide association study (GWAS) of IDC in soybean, which located a previously reported locus in addition to a novel locus involved in IDC resistance [8]. CNN analysis was used to detect the presence of northern leaf blight lesions in corn plants in the field with 96.7% accuracy [80]. A complex dataset composed of tomato leaf and plant images with cluttered backgrounds often containing other plant parts, fruit, or greenhouse structures was employed to develop a multilevel approach applying region-based CNNs (R-CNNs). It generated bounding boxes around symptomatic areas in each image [81], followed by a CNN filter bank to reduce 'false positives', resulting in a 96% recognition rate of disease, insect, and abiotic stress symptoms [82].

A ‘mini-plot’ hyperspectral imaging system was designed to image the early stages of disease infection in barley canopies under more controlled environmental conditions [83]. ML methods include simplex volume maximization (SiVM), which reduced data size, and SVM that works as a classifier for healthy tissue, disease symptoms, and background. These models allowed powdery mildew disease phenotyping of six barley cultivars and classification into low, medium, or high severity categories [83]. In maize, close-range hyperspectral imaging under controlled conditions was employed to detect early drought stress in vegetative stage plants [84]. In wild tomato, quadcopters mounted with RGB and multispectral cameras were used to image accessions, followed by morphometric and spectral analysis to identify the highest-performing accessions in response to salinity stress [85].

The response of plant height can be related to yield and stress tolerance, and has been evaluated using stereo RGB imaging in wheat to determine the nitrogen response of different cultivars [86]. 3D laser scanning and LIDAR (light detection and ranging) have been used to reconstruct 3D point clouds of crop canopies. 3D laser scanning has previously analyzed maize, soybean, and wheat canopy height growth under field conditions [87] and a small-scale analysis of barley organ responses to irrigation [88]. The 3D-based canopy architecture is also important for measuring stress responses in the field, and LIDAR was able to unravel peanut canopy characteristics in the field [89]. Eggplant seedling growth was examined in potted pots with high-resolution optical probe-based scanning LIDAR, which also monitored 3D shape changes of potted tomato plant leaves in response to water stress [90]. A structure from motion (SfM) technique with segmentation reconstructed a 3D model of 20 sugar beet genotypes at three stages, and included plant height, total leaf area, canopy area, total leaf area, and leaf length [91]. Other disease traits that will benefit from 3D phenotyping include those that cause stunting and leaf wrinkling [92]. Implementing 3D technologies with appropriate data analytics adds an additional dimension to phenotyping, enabling data collection of plant structural response to stresses.

Field-Scale Foliar Stress Phenotyping

Field researchers including plant breeders, agronomists, pathologists, and entomologists design large field plot tests to closely resemble farmer field conditions. These experiments include testing the efficacy of resistance genes, fungicide and herbicide trials for testing stress management strategies, and cultural practices. These tests or fields depend on higher-throughput phenotyping systems to efficiently collect phenotypic data from hundreds to thousands or more areas of production fields, depending on the research platform size, in screening or scouting for plant stress severity. Unmanned aerial systems have emerged as a promising strategy for the implementation of imaging systems in the field for HTP of foliar stress [93] and agronomic research [94–96]. The use of rotary and fixed-wing UAS to monitor water stress, crop nutrient status, crop diseases, crop pests, and weeds has been reviewed [93].

In soybean, IDC imaging in the field was accomplished through UAS followed by ML-directed classification on the 1–5 severity scale commonly used for IDC. This method has 77% accuracy [97]. Weed pressure quantification was realized through UAS-based imaging including weed identification, counting, and mapping [98]. Commonly mounted sensing instruments for UAS agronomic and/or disease monitoring include RGB cameras [99], multispectral cameras [85], hyperspectral cameras, and thermal cameras [100]. Although progress has been made in UAS implementation for stress detection [101] and stress level categorization [102], more work will be necessary to drive stress severity quantification for utilization in genetic studies, germplasm screening, and in scouting decision aid tools. A potential explanation for this paucity of studies may be the negative correlation between efficient flight time and the high image resolution that is needed for examining small foliar changes, as well as difficulties associated with the adaptation of ML methods for on-board analytics in the field. Improvements in imaging technology, including higher-resolution cameras and UAS hardware that enable higher (and longer) altitude flights with sub-centimeter resolution, will improve the accuracy of UAS-based HTP of foliar stresses especially in the areas of plant and weed identification, stress detection and severity quantification, and data translation into actionable items to enable decision support.

Developing an Overarching Strategy for Plant Stress Phenotyping

Challenges in the visual assessment of stress, such as rater reliability, experience level variation, training requirements, fatigue, and inherent bias, have prompted calls for machine-based stress phenotyping systems [26]. As previously explained, an overarching foliar stress phenotyping system should build on the successes in leaf, canopy, and field phenotyping. Future strategies should improve (i) objectivity, (ii) scalability, and (iii) versatility of current plant stress rating strategies. Furthermore, to cultivate future advances and put research resources to their most effective use, new phenotyping strategies should also incorporate the FAIR guiding principles – that all research products should be findable, accessible, interoperable, and reusable for both machines and people [103], and improve the distribution of research findings. Following FAIR principles should help in setting standards for creating image-based stress datasets. Specifically, for the plant sciences and phenotyping community, this implies that each dataset (stemming from unique studies) is accompanied by a framework that simplifies data retrieval as researchers add unique and persistent identifiers to the data; for example, digital object identifiers (DOI), institutional repository case IDs if protected under licensing agreements, and secure and active URLs for archiving on the web. One of the other major requirements under FAIR principles is the addition of metadata because descriptors provide consistency in interpretation and application by the broader community. It will also facilitate data retrieval with relevant search terms in databases. To realize the value of shared data, particularly for data interrogation and merging by researchers, data/metadata should follow predetermined crop community standards regarding structure, vocabulary, and ontology. The crop protection (pathologists and entomologists) and plant breeding communities should develop standards for plant disease image data, with a coherent structure and approved vocabularies and ontologies for automated and seamless integration across domains. It will reduce redundancies and lead to cost savings. For example, the current trend in plant sciences is to share research data from published work through supplementary files or data repositories (e.g., in the case of soybean, genomic data are shared with soybean community using web resources such as SoyBase (https://soybase.org/). Data sharing through repositories allows researchers to demonstrate ownership in generating data. The goal is to strategically guide development of stress phenotyping methods to improve the quality of data and the distribution of current phenotyping systems to move from more disjointed approaches to a unified approach.

Accurate phenotyping is necessary for improving the inheritance values of valuable traits needed for stress resistance [104]. Visual rating methods assisted by SADs or iPhone apps decreases inter-/intra-rater variation, but there still can be substantial subjectivity and variability among raters (lack of training, experience level, or field fatigue). Image-based phenotyping strategies can increase objectivity because these techniques apply objective rules to phenotyping and are less likely to be affected by human error. Furthermore, unsupervised learning decreases the amount of subjectivity introduced by human error during the expert labeling that is necessary for supervised and semi-supervised learning approaches. Visual ratings are limited to the amount of information that is visible to the human eye. Deeper phenotyping can be achieved through the use of remote sensing, such as with hyperspectral and thermal cameras that examine wavelengths beyond human vision, and that have identified deep tissue disease symptoms as well as detecting disease signatures before visual symptoms [40,105]. These sensors also enable the detection of smaller subtle changes in disease symptoms at the micro- and macroscopic levels, such as in the early stages of infection [106].

Developing stress phenotyping methods that can facilitate variable scales, or the scalability, of operations is also important. Rating current stress is as limiting as it is laborious and time-consuming owing to the nature of human-dependent ratings. New raters need to be trained and retrained continuously so that they can reach peak performance. Even then, there is a limit to the time and funding that can be allocated to the collection of a thorough dataset. Depending on objectives of a study, phenotyping may be needed at individual leaf, plant canopy, plot, or field scales. Scalability of phenotyping methods can improve robustness of phenotyping at a variety of scales to capture desired information efficiently (at spatial and temporal scales).

Versatility is an important consideration in developing any impactful new technology, and this is necessary for estimating plant stress severity (similarly to next-generation DNA sequencing methodologies) to allow wider applicability across crop species and stress types. Current methods lack comparable versatility to genotyping platforms because they cannot directly transfer to other crop/disease complexes. The variation in disease symptoms, composite plant morphology, or leaf structure among different genotypes of the same crops may lead to reduced accuracy and precision in disease severity estimates [45]. Theoretically, there are multiple ways to estimate or measure disease severity percentages. For example, symptoms could be localized to one area or be evenly distributed throughout the plant structure, or could be present as many small lesions or several large lesions; and these cause issues with accuracy in visual ratings [15]. To attain the goal of versatility, a plant severity measurement system must be developed and tested for each target disease and crop type. Compared with manual ratings, information extracted by ML is flexible and useful to solve phenotyping issues related to crop stress. Novel and automated methods with increased sensitivity, specificity, and reliability would improve disease detection beyond visual ratings [10]. Interlacing these new methods and sensors into current disease rating strategies will allow more objective and robust ratings across a wide scale environment, while also promoting versatility in program goals or disease/crop system monitoring. For efficient and effective implementation at large scales, improvements in phenotyping platforms and sensors will be needed.

ML Concepts that Simplify Transition to Practice: Transfer Learning

Advances in ML techniques offer methods that can improve the speed of adoption of ML for agricultural phenotyping. **Transfer learning** (Figure 2) is a technique that aims to transfer knowledge gained from solving a task in the source domain to solving an unseen task in the target domain where the amount of data available for training is scarce [107]. In transfer learning, a model is generally pretrained using an abundant amount of training data to solve a task in the source domain, and is then fine-tuned for solving the target domain task. Because DL models generally require a large dataset of images to learn from scratch, using plant images to fine-tune the pretrained model architecture can significantly improve model performance on the target task. In supervised pretraining, labels of source domain data are used for training the model, whereas they are not used in the unsupervised pretraining method. It has recently been shown that pretraining a large model on large amounts of data in the source domain can reduce the amount of training data needed for fine-tuning the model to solve the target domain task [108,109]. The data efficiency for solving the target task also depends on the amount of similarity between the source and target domains [110]. For example, it is more challenging to transfer knowledge from representations learned using RGB images to a target task using hyperspectral

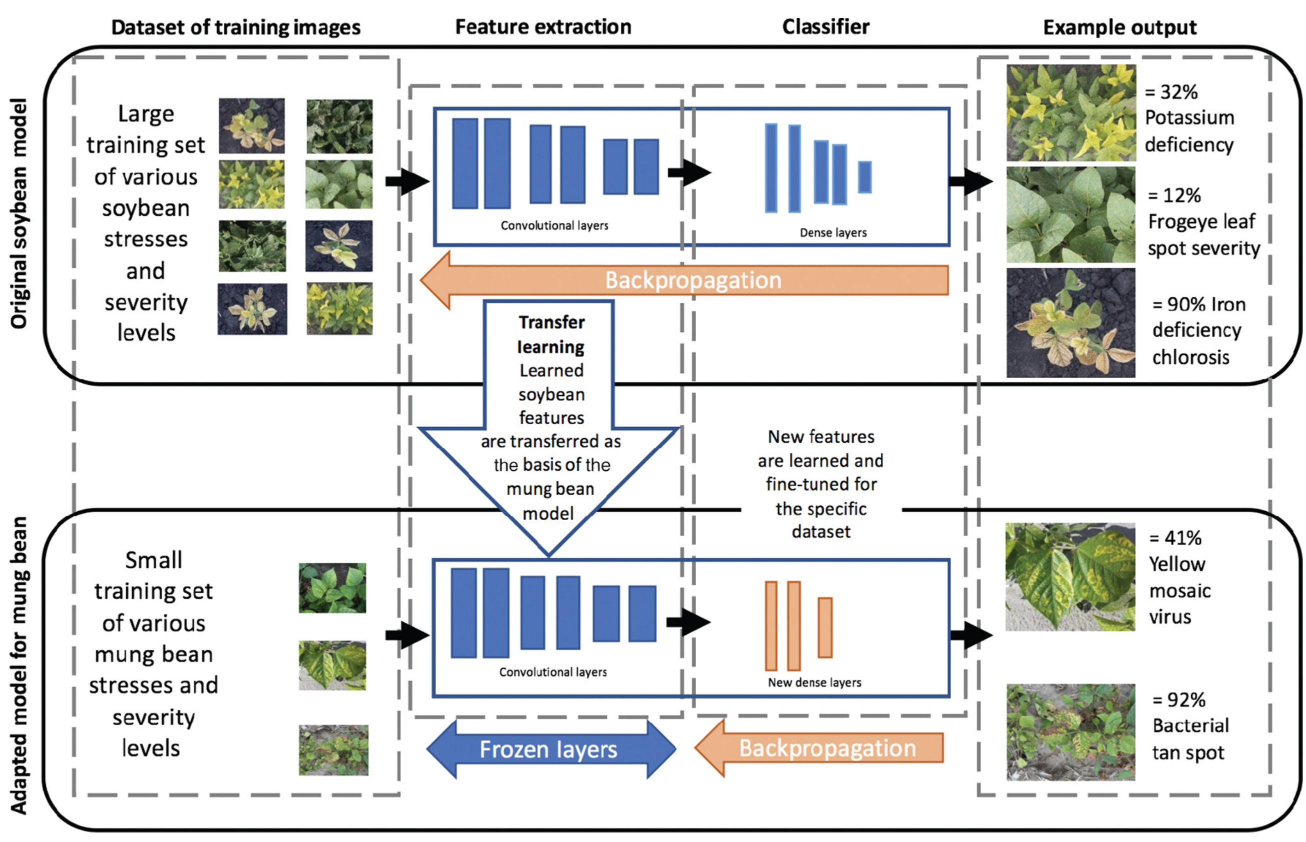

Figure 2. Transfer Learning for Optimization of the Learning Process in Deep Learning (DL) for Plant Stress Phenotyping. The original DL model is trained on soybean disease images and, by using transfer learning, the soybean disease classification model is fine-tuned to classify mung bean disease images.

images from UAV or satellite. The transferability of representations is also challenging if there is minimal overlap in disease symptomatology between the source and target domain. For example, rusts in monocots and dicots have dissimilar disease phenotypes. Thus, more training data will be necessary to solve the target task owing to dissimilarity between the source and target domains. However, disease symptoms in the same pathogen genus, and on occasions across genera, may have some common minimum features that can be utilized during transfer learning. Transfer learning techniques allow us to transfer the representations learned by a DL model on one disease or crop system to another, speeding up the time required for training, and smoothing the transition into new crop or disease systems. It is important to realize that pretrained architectures (using non-plant images) can serve as the backbone for the DL models used in the target task (using plant images).

We consider transfer learning to broadly encompass not only the conventional notion of network pretraining but also the newer notion of domain adaptation (both in-domain and cross-domain adaptation). This includes unique approaches for self and semi-supervised learning that deal with training models for 'learning to learn' such that they are trained with data from one domain. Then, they can be (re)trained not only on the last few layers but on the full network for a completely new domain. Transfer learning can provide a better, cheaper, and faster solution to the issue of limited training data availability. Researchers working on apple black rot evaluated the performance of trained shallow networks (trained from scratch) and deep models (fine-tuned by transfer learning) to diagnose the severity of the disease, and reported that the deep VGG16 model trained with transfer learning was the best model, with an overall accuracy of 90.4%. Simple features such as edges and blobs learned in the initial layers of the model can be transferred intact between different disease DL models, thus leveraging pre-existing learning networks while avoiding the majority of expensive data labeling [107]. This phenomenon can be used efficiently through transfer learning of already available trained networks [31,72,75,111], and transfer learning has been used in a wide range of experiments [67,69,112,113].

Researchers compared AlexNet and GoogLeNet architectures based on a training method for predicting correct crop–disease pairs given 38 possible classes. Among the two architectures and various methods of learning used in the study, GoogLeNet-based transfer learning yielded the best results, achieving an accuracy of 99.34% [72]. However, the benefits of transfer learning can be themselves limited. For example, Barbedo examined the impact of data size on the effectiveness of DL and transfer learning for plant disease classification using a 1381 image dataset comprising 12 plant species and 56 classes (diseases) [111]. Separate CNNs were trained to study the effect of image backgrounds. Classification accuracies with the original images varied from 65% (common bean, with 64 images representing five classes) to 100% (cotton, with 95 images representing three classes). In their experiment the limited size of the dataset used for fine-tuning restricted the scope of transfer learning for predicting the diverse set of diseases [111]. In addition, background was also an important factor in accuracy leading to the conclusion that, for practical purposes, a wide variety of image backgrounds will need to be included to dilute its effect in analysis under real circumstances. Therefore, the application of transfer learning may still require larger datasets under real conditions for higher accuracies [114].

ML Concepts That Simplify Transition to Practice: AL

Typically, ML models are trained using a passive learning philosophy. This involves gathering a large amount of data randomly sampled from the data distribution and using this large dataset to train a model to perform identification, classification, and quantification tasks [115]. Data annotation or labeling is a crucial stage in supervised ML. Acquiring high-quality labeled data is a developmental barrier to building a complex DL model. In plant sciences, expert annotators are needed. The task of annotation requires domain-specific knowledge, especially when confounding symptoms for various plant stresses are present. Data annotation by experts with domain-specific knowledge is a tedious and expensive task. AL is a smart data annotation strategy that can partly automate the process of data annotation and reduce the need for substantial human engagement by identifying the most informative and non-redundant samples for training the model [28]. The AL approach works by iteratively increasing the size of judiciously selected labeled data to achieve comparable or greater performance than fully supervised ML models, and can be trained at a fraction of the cost or time that it takes to label all the data. The process of weakly supervised DL, as demonstrated in a study of sorghum head detection [36], involves carefully selecting instances based upon the collected data rather than randomly selecting those instances. Such AL-based ML methods have potential for plant stress identification, classification, and quantification. Further development of such approaches will become mainstream in the future because they require substantially less data for training a model and can use a pretrained DL model. Therefore, the choice of how much data to use, or alternatively how much performance is desired from the model, relies on a resource management decision and on the sample diversity in the collected dataset.

Challenges in Applying ML to Image-Based Plant Stress Phenotyping

The adoption of new methods and technologies such as automated phenotyping is not without challenges and barriers to implementation, including barriers to entry and user aversion [116]. Current challenges can be organized into three main categories: data collection, model training, and model transferability.

High-throughput phenotypic data collected by HTP platforms such as UAS, field robots, or tractor-mounted equipment demonstrates a higher level of occlusion and background noise compared with destructive or laboratory-based image datasets [95,117,118]. The nature of field implementation of automatic phenotyping adds complex variations into data quality, including lighting variations due to variable cloud cover, angle, and intensity of sunlight, or wind intensity, that result in algorithm obstacles [119–122]. Furthermore, advanced sensors such as hyperspectral cameras and LIDAR contribute additional information, but also additional cost in hardware, data analytics, and the interpretation of information [123].

Ensuring the value of training data to improve algorithm performance is an extremely important and challenging aspect of working with ML [124]. The development of the PlantVillage database that contained ~54 306 images of 14 crop species and 26 diseases created new opportunities to meet some of the research needs (number of diseases, crops, severity of expression, stages of disease infection, etc.), including the need for a publicly available, open-source, shared database of annotated plant stresses at individual leaf scale. PlantVillage data have also been overly explored, and new sources of data will be necessary for the development of robust models. Other open-source databases are also available, such as the maize northern leaf blight (NLB) disease database [125], which comprises 18 222 digital images of maize leaves in field, either taken manually, mounted on a boom, or by UAS. Around 105 705 NLB lesions were annotated by human experts, making this the largest publicly available image set annotated for a plant disease. Selvaraj *et al.* trained a DCNN model for banana disease and pest detection with >90% accuracy using ~18 000 images collected from farmers' fields in Africa, Latin America, and India [126]. The authors have made all images from their study accessible to the research community (https://pestdisplace.org/). Another dataset developed by Embrapa is known as the plant disease database (PDDB), consisting of 2326 images of 171 diseases and other disorders affecting 21 plant species [127]. To increase the size of the database, images were subdivided to increase the number of images to 46 513.

Annotated image databases containing images sourced from collaborating institutions fulfill the need for large datasets for effectively implementing DL and transfer learning in disease classification and, by extension, severity assessment [127]. The use of hyperspectral sensors will also help in the implementation of DL because they generate a large amount of information and high-dimensional datasets from a limited number of samples [31]. The development of a single database for annotated plant stress images also supports the FAIR principles of accessibility and reusability. Great care must also be taken to ensure that the training data exemplify the variability and complexity the algorithm may encounter in practice to ensure high performance accuracy. Plant stress image data collected in a single year, at a single location, or on a single plant variety or crop stage will not span the variability of trait expression, resulting in an algorithm that cannot be applied to real-life scenarios in different settings [128]. Therefore, proper early planning of experiments, sensors, resolution, and end-use of images, ideally in a disciplined, collaborative manner, will be essential.

Finally, advances in AL and transfer learning can further augment ML models for new objectives and plant stresses (disease, drought, flooding, salinity, temperature, nutrient, pest, weed stresses). However, one of the other major community resource innovations will be an annotated plant stress dataset such as ImageNet, which consists of ~14 million images. The availability of such datasets for training DL layers and models on plant images at individual leaf, canopy, and plot scale increases the possibility of obtaining high-performance ML models for practical application in farmers' fields or routine use in research settings or breeding programs [129]. Transfer learning using pretrained networks such as ImageNet, AlexNet, ResNet still poses challenges because they are non-plant image-based pretrained DL models. Similarly, a model trained to identify, classify, or quantify stress at plot/field scale is not directly applicable at the individual leaf scale, and vice versa. Stress symptoms in images collected in natural field settings are different from those collected under artificial settings (e.g., in greenhouses) to create disease or insect pressure. Therefore, strategic collection of training data is essential to allow an algorithm to be successfully deployed. The development of very large and diverse plant stress datasets for DL model training can be achieved by combining previous large crowd-sourced datasets. Given the availability of annotated image databases, and more importantly of data from multispectral/hyperspectral and other sensors, complemented with AL approaches and/or transfer learning (with plant stress images), it will be possible to use AI for real-time plant stress identification, classification, and quantification at field scale. Statistical tools leveraging ML techniques are rapidly advancing, and will achieve spectral image super-resolution, specifically working with spectral images at varying scales of spatial resolution, as well as linking spectral and trichromatic (RGB) images. This will enable cost-effective disease phenotyping [130,131].

Concluding Remarks and Future Perspectives

Plant stress severity phenotyping is an important parameter for assessing potential crop losses due to various biotic and abiotic stresses. It can be employed to identify superior disease-resistant and stress-tolerant genotypes and to evaluate disease management decisions. Current methods for stress severity phenotyping are deployed at various scales, such as exact counts of lesion numbers or the number of plants affected, or estimates of the severity or surface area affected by a particular stress at canopy and field levels. ML is a promising solution for improving the speed, accuracy, reliability, and scalability of image-based disease phenotyping while allowing flexibility for highly variable program goals at research plots and eventually farmer's fields. These include improved disease rating data quality by decreasing human error, and to some extent inter- and intra-rater variation, among other issues. Both ML and DL can be seamlessly integrated into data acquisition, data preprocessing, and data analytics for real-time HTP of

Outstanding Questions

How can automated disease rating scales be designed that can evaluate crop varieties for stress resistance, herbicide efficacy, and pest management at the whole plant, plot, or field level?

How can automated disease rating scales be deployed in breeding and research programs to enable HTP using advanced sensors?

How can multiple programs with varying goals work together to aid ML development given the need for larger training datasets for handling wider objectives?

DL techniques require larger training datasets. Can we develop intelligent strategies to reduce the need for larger training datasets?

How can the rate of dissemination of DL algorithms be improved across crops, stresses, and program goals to aid multiple research programs and producers by developing useful evaluation tools?

Box 1. Recommendations for Future Analysis

(i) The amount of data required to train ML models depends on the complexity of the problem and the complexity of the learning algorithm; therefore, for wider applicability, the training data should be continuously updated using techniques such as AL to reflect the complexity of stress symptoms for the crop in question.

(ii) To train robust ML models for practical application in decision support, the plant community needs universal datasets for different stresses. The datasets should include realistic and potentially degraded sensing environments (e.g., cloudy, low light, fog, saturated lighting) to ensure robust in-field performance of ML. Collaboration will be required between different disciplines to aid developers of ML while enlarging training datasets to increase that capability and versatility of the algorithms. Such an approach will enable development pipelines that are capable of handling diverse objectives.

(iii) Higher-resolution sensors such as hyperspectral snapshot cameras and thermal cameras would promote leaf-level quality phenotyping at the rate and scale of field phenotyping.

(iv) Higher-quality DL models will benefit from expanded training datasets. Improvement in sensors, data quality, image capture, and the efficiency of HTP will increase the amount of data and the rate of improvement of DL models, and, together with analytics, are essential components of this pipeline.

(v) Translation of rapidly developing strategies such as (cross) domain adaptation [132] and self-supervised and semi-supervised learning to the plant science community can help to create generalizable phenotyping workflows (instead of the current trend towards crop/species/stress-specific ML tools) that will aid multiple research programs and producers with the development of useful evaluation tools.

(vi) Current ML-based strategies focus on a single disease or stress located on a leaf or canopy, but, in real-world situations, multiple diseases and stresses may appear on a single leaf or on a single plant canopy. ML platforms must be robust and flexible, and be able to differentiate between multiple disease symptoms on a single leaf or on the same plant canopy. The training dataset should contain multi-year, multi-location, and diverse symptom images of plant stresses.

(vii) In breeding and research programs, larger trials are necessary for the evaluation of new crop varieties, herbicide efficacy, and pest management strategies. Systems that can evaluate crops at the whole plant, plot, or field level will enable full use of the advantages of HTP and advanced sensors.

(viii) An open-source online repository should be organized consisting of plant stress datasets for each crop, with detailed best practice guidelines for data collection at various phenotyping scales (individual leaf, canopy, and field scale). If stress is of national importance, then state-wide data collection is necessary. For stresses of international importance, for example rust epidemics in primary food crops such as wheat, then nation-wide data collection will be useful for capturing the entire spectrum of disease or stress expression in question.

(ix) The collected image-based stress dataset should meet FAIR data principles (findability, accessibility, interoperability, and reusability) [103].

(x) There is a need to create a scalable cyberinfrastructure for data collection, data curation, data storage, and data analysis.

plant traits in the field. Ongoing efforts are needed in the development and application of ML methods in the quantification and prediction of disease severity to complement advances in sensor and phenotyping platforms. These efforts will complement advances in imaging technology for high-resolution and high-dimensional data collection. Machine-augmented plant stress phenotyping will provide state-of-the-art solutions to farmers, plant breeders, pathologists, and plant scientists (see Outstanding Questions). We conclude with a list of recommendations for future studies (Box 1).

Acknowledgments

This work was supported by US Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) Food and Agriculture Cyberinformatics Tools (FACT) award 2019-67021-29938 (A.S., B.G., S.S), Iowa Soybean Research Center (A.S.), Plant Science Institute Fellowship (B.G., S.S.), National Science Foundation (NSF)/USDA NIFA award 2017-67007-26151 (S.S., B.G., A.S.), and USDA Current Research Information System (CRIS) IOW04314 project (A.S.). We thank Meg Grice for editing the article.

References

1. Strange, R.N. and Scott, P.R. (2005) Plant disease: a threat to global food security. *Annu. Rev. Phytopathol.* 43, 83–116
2. Anderson, P.K. *et al.* (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol. Evol.* 19, 535–544
3. Evans, N. *et al.* (2008) Range and severity of a plant disease increased by global warming. *J. R. Soc. Interface* 5, 525–531
4. Deutsch, C.A. *et al.* (2018) Increase in crop losses to insect pests in a warming climate. *Science* 361, 916–919

5. Bock, C.H. *et al.* (2010) Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Crit. Rev. Plant Sci.* 29, 59–107
6. Ghosal, S. *et al.* (2018) An explainable deep machine vision framework for plant stress phenotyping. *Proc. Natl. Acad. Sci. U. S. A.* 115, 4613–4618
7. Naik, H.S. *et al.* (2017) A real-time phenotyping framework using machine learning for plant stress severity rating in soybean. *Plant Methods* 13, 23
8. Zhang, J. *et al.* (2017) Computer vision and machine learning for robust phenotyping in genome-wide studies. *Sci. Rep.* 7, 44048
9. Fiorani, F. and Schurr, U. (2013) Future scenarios for plant phenotyping. *Annu. Rev. Plant Biol.* 64, 267–291
10. Mahlein, A.K. (2016) Plant disease detection by imaging sensors – parallels and specific demands for precision agriculture and plant phenotyping. *Plant Dis.* 100, 241–251
11. Araus, J.L. *et al.* (2018) Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.* 23, 451–466
12. Krupinsky, J.M. *et al.* (2002) Managing plant disease risk in diversified cropping systems. *Agron. J.* 94, 198
13. Kranz, J. and Rotem, J. (2012) *Experimental Techniques in Plant Disease Epidemiology*, Springer Science and Business Media
14. Nutter, F.W. *et al.* (1991) Disease assessment terms and concepts. *Plant Dis.* 75, 1187–1188
15. Madden, L.V. *et al.* (2007) *Study of Plant Disease Epidemics*, American Phytopathological Society
16. Bock, C.H. *et al.* (2016) Assessing disease severity: accuracy and reliability of rater estimates in relation to number of diagrams in a standard area diagram set. *Plant Pathol.* 65, 261
17. Nutter, J.F.W. (1993) Assessing the accuracy, intra-rater repeatability, and inter-rater reliability of disease assessment systems. *Phytopathology* 83, 806
18. Bock, C.H. *et al.* (2015) Disease severity estimates-effects of rater accuracy and assessment methods for comparing treatments. *Plant Dis.* 99, 1104
19. Chiang, K.-S. *et al.* (2016) Plant disease severity assessment-how rater bias, assessment method, and experimental design affect hypothesis testing and resource use efficiency. *Phytopathology* 106, 1451
20. Akintayo, A. *et al.* (2018) A deep learning framework to discern and count microscopic nematode eggs. *Sci. Rep.* 8, 9145
21. Newton, A.C. and Hackett, C.A. (1994) Subjective components of mildew assessment on spring barley. *Eur. J. Plant Pathol.* 100, 395–412
22. Sherwood, R.T. *et al.* (1983) Illusions in visual assessment of *Stagonospora* leaf spot of orchardgrass. *Phytopathology* 73, 173–177
23. Yadav, N.V.S. *et al.* (2013) Development and validation of standard area diagrams to aid assessment of pecan scab symptoms on fruit. *Plant Pathol.* 62, 325–335
24. Bock, C.H. *et al.* (2008) Visual rating and the use of image analysis for assessing different symptoms of citrus canker on grapefruit leaves. *Plant Dis.* 92, 530
25. Nita, M. *et al.* (2003) Reliability and accuracy of visual estimation of phomopsis leaf blight of strawberry. *Phytopathology* 93, 995
26. Singh, A. *et al.* (2016) Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.* 21, 110
27. Rumpf, T. *et al.* (2010) Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Comput. Electron. Agric.* 74, 91–99
28. Settles, B. (2009) *Active Learning Literature Survey*, University of Wisconsin-Madison Department of Computer Sciences
29. Zhu, X. and Goldberg, A.B. (2009) Introduction to semi-supervised learning. In *Synthesis Lectures on Artificial Intelligence and Machine Learning* (Vol. 3) (Brachman, R. *et al.*, eds), pp. 1–130, Morgan & Claypool Publishers
30. Jing, L. and Tian, Y. (2020) Self-supervised visual feature learning with deep neural networks: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* Published online May 4, 2020. https://doi.org/10.1109/TPAMI.2020.2992393
31. Singh, A.K. *et al.* (2018) Deep learning for plant stress phenotyping: trends and future perspectives. *Trends Plant Sci.* 23, 883–898
32. Mitchell, T.M. (1997) *Machine Learning*, McGraw-Hill Higher Education
33. Ma, C. *et al.* (2014) Machine learning for big data analytics in plants. *Trends Plant Sci.* 19, 798–808
34. Amorim, W.P. *et al.* (2019) Semi-supervised learning with convolutional neural networks for UAV images automatic recognition. *Comput. Electron. Agric.* 164, 104932
35. Qin, F. *et al.* (2016) Identification of alfalfa leaf diseases using image recognition technology. *PLoS ONE* 11, e0168274
36. Ghosal, S. *et al.* (2019) A weakly supervised deep learning framework for sorghum head detection and counting. *Plant Phenomics* 2019, 1525874
37. Bah, M.D. *et al.* (2018) Deep learning with unsupervised data labeling for weed detection in line crops in UAV images. *Remote Sens.* 10, 1690
38. Pardede, H.F. *et al.* (2018) Unsupervised convolutional autoencoder-based feature learning for automatic detection of plant diseases. In *2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, pp. 158–162, IEEE
39. Lecun, Y. *et al.* (2015) Deep learning. *Nature* 521, 436–444
40. Nagasubramanian, K. *et al.* (2019) Plant disease identification using explainable 3D deep learning on hyperspectral images. *Plant Methods* 15, 98
41. Yamamoto, K. *et al.* (2017) Super-resolution of plant disease images for the acceleration of image-based phenotyping and vigor diagnosis in agriculture. *Sensors* 17, 2557
42. Molina, P. *et al.* (2017) First results of a tandem terrestrial-unmanned aerial mapKITE system with kinematic ground control points for corridor mapping. *Remote Sens.* 9, 60
43. O'Toole, M. *et al.* (2018) Confocal non-line-of-sight imaging based on the light-cone transform. *Nature* 555, 338–341
44. Corrêa, F.M. *et al.* (2009) Comparison of three diagrammatic keys for the quantification of late blight in tomato leaves. *Plant Pathol.* 58, 1128–1133
45. Costa Lage, D.A. *et al.* (2015) Standard area diagrams for assessment of powdery mildew severity on tomato leaves and leaflets. *Crop Prot.* 67, 26–34
46. James, W.C. (1971) An illustrated series of assessment keys for plant diseases, their preparation and usage. *Can. Plant Dis. Surv.* 51, 39–65
47. Librelon, S. *et al.* (2015) Diagrammatic scale to evaluate angular leaf spot severity in primary leaves of common bean. *J. Australas. Plant Pathol. Soc.* 44, 385–395
48. Large, E.C. (1966) Measuring plant disease. *Annu. Rev. Phytopathol.* 4, 9–26
49. Duan, J. *et al.* (2015) Development and validation of a standard area diagram set to aid estimation of bacterial spot severity on tomato leaves. *Eur. J. Plant Pathol.* 142, 665–675
50. Duarte, H. *et al.* (2013) Development and validation of a set of standard area diagrams to estimate severity of potato early blight. *Eur. J. Plant Pathol.* 137, 249–257
51. Capucho, A.S. *et al.* (2010) Influence of leaf position that correspond to whole plant severity and diagrammatic scale for white spot of corn. *Crop Prot.* 29, 1015–1020
52. Godoy, C.V. *et al.* (1997) Diagrammatic scales for bean diseases: development and validation. *J. Plant Dis. Protect.* 104, 336–345
53. Maciel, J.L.N. *et al.* (2013) Diagrammatic scale for the assessment of blast on wheat spikes. *Summa Phytopathol.* 39, 162–166
54. Spolti, P. *et al.* (2011) Improving sooty blotch and flyspeck severity estimation on apple fruit with the aid of standard area diagrams. *Eur. J. Plant Pathol.* 129, 21–29
55. Del Ponte, E.M. *et al.* (2017) Standard area diagrams for aiding severity estimation: scientometrics, pathosystems, and methodological trends in the last 25 years. *Phytopathology* 107, 1161
56. Rios, J. *et al.* (2013) Development and validation of a standard area diagram set to assess blast severity on wheat leaves. *Eur. J. Plant Pathol.* 136, 603–611
57. González-Domínguez, E. *et al.* (2014) Development and validation of a standard area diagram set to aid assessment of severity of loquat scab on fruit. *Eur. J. Plant Pathol.* 139, 413–422

58. Debona, D. *et al.* (2015) A set of standard area diagrams to assess severity of frogeye leaf spot on soybean. *Eur. J. Plant Pathol.* 142, 603–614
59. Lamari, L. (2002) *Assess: Image Analysis Software for Plant Disease Quantification*, APS Press
60. Barbedo, J. (2014) An automatic method to detect and measure leaf disease symptoms using digital image processing. *Plant Dis.* 98, 1709–1716
61. Schneider, C.A. *et al.* (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675
62. Stewart, E.L. and McDonald, B.A. (2014) Measuring quantitative virulence in the wheat pathogen *Zymoseptoria tritici* using high-throughput automated image analysis. *Phytopathology* 104, 985
63. Pethybridge, S.J. and Nelson, S.C. (2015) Leaf Doctor: a new portable application for quantifying plant disease severity. *Plant Dis.* 99, 1310
64. Pethybridge, S.J. and Nelson, S.C. (2018) Estimate, a new iPad application for assessment of plant disease severity using photographic standard area diagrams. *Plant Dis.* 102, 276
65. Bardsley, S.J. and Ngugi, H.K. (2013) Reliability and accuracy of visual methods to quantify severity of foliar bacterial spot symptoms on peach and nectarine. *Plant Pathol.* 62, 460–474
66. Hughes, D.P. and Salathe, M. (2015) *An open access repository of images on plant health to enable the development of mobile disease diagnostics*.
67. Ferentinos, K.P. (2018) Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145, 311–318
68. Atabay, H.A. (2017) Deep residual learning for tomato plant leaf disease identification. *J. Theor. Appl. Inf. Technol.* 95, 6800–6808
69. Brahimi, M. *et al.* (2017) Deep learning for tomato diseases: classification and symptoms visualization. *Appl. Artif. Intell.* 31, 299–315
70. Zhang, K. *et al.* (2018) Can deep learning identify tomato leaf disease? *Adv. Multimedia* 2018, 6710865
71. Zhang, S. *et al.* (2019) Three-channel convolutional neural networks for vegetable leaf disease recognition. *Cogn. Syst. Res.* 53, 31–41
72. Mohanty, S. *et al.* (2016) Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7, 1419
73. Too, E.C. *et al.* (2019) A comparative study of fine-tuning deep learning models for plant disease identification. *Comput. Electron. Agric.* 161, 272–279
74. Boulent, J. *et al.* (2019) Convolutional neural networks for the automatic identification of plant diseases. *Front. Plant Sci.* 10, 941
75. Wang, G. *et al.* (2017) Automatic image-based plant disease severity estimation using deep learning. *Computat. Intell. Neurosci.* 2017, 2917536
76. Lin, K. *et al.* (2019) Deep learning-based segmentation and quantification of cucumber powdery mildew using convolutional neural network. *Front. Plant Sci.* 10, 155
77. Charlson, D. *et al.* (2009) Polygenic inheritance of canopy wilting in soybean [*Glycine max* (L.) Merr.]. *Theor. Appl. Genet.* 119, 587–594
78. King, C.A. *et al.* (2009) Differential wilting among soybean genotypes in response to water deficit. *Crop Sci.* 49, 290–298
79. Atoum, Y. *et al.* (2016) On developing and enhancing plant-level disease rating systems in real fields. *Pattern Recogn.* 53, 287–299
80. DeChant, C. *et al.* (2017) Automated Identification of northern leaf blight-infected maize plants from field imagery using deep learning. *Phytopathology* 107, 1426
81. Fuentes, A. *et al.* (2017) A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors* 17, 2022
82. Fuentes, A.F. *et al.* (2018) High-performance deep neural network-based tomato plant diseases and pests diagnosis system with refinement filter bank. *Front. Plant Sci.* 9, 1162
83. Thomas, S. *et al.* (2018) Quantitative assessment of disease severity and rating of barley cultivars based on hyperspectral imaging in a non-invasive, automated phenotyping platform. *Plant Methods* 14, 45
84. Asaari, M.S.M. *et al.* (2019) Analysis of hyperspectral images for detection of drought stress and recovery in maize plants in a high-throughput phenotyping platform. *Comput. Electron. Agric.* 162, 749–758
85. Johansen, K. *et al.* (2019) Unmanned aerial vehicle-based phenotyping using morphometric and spectral analysis can quantify responses of wild tomato plants to salinity stress. *Front. Plant Sci.* 10, 370
86. Cai, J. *et al.* (2018) Land-based crop phenotyping by image analysis: accurate estimation of canopy height distributions using stereo images. *PLoS ONE* 13, e0196671
87. Friedli, M. *et al.* (2016) Terrestrial 3D laser scanning to track the increase in canopy height of both monocot and dicot crop species under field conditions. *Plant Methods* 12, 9
88. Paulus, S. *et al.* (2014) Automated analysis of barley organs using 3D laser scanning: an approach for high throughput phenotyping. *Sensors* 14, 12670
89. Yuan, H. *et al.* (2019) Development of a peanut canopy measurement system using a ground-based LiDAR sensor. *Front. Plant Sci.* 10, 203
90. Omasa, K. *et al.* (2007) 3D lidar imaging for detecting and understanding plant responses and canopy structure. *J. Exp. Bot.* 58, 881–898
91. Xiao, S. *et al.* (2020) Image-based dynamic quantification of aboveground structure of sugar beet in field. *Remote Sens.* 12, 269
92. Hartman, G.L. *et al.* (2015) *Compendium of Soybean Diseases and Pests* (5th edn), American Phytopathological Society
93. Barbedo, J.G.A. (2019) A review on the use of unmanned aerial vehicles and imaging sensors for monitoring and assessing plant stresses. *Drones* 3, 40
94. Maes, W.H. and Steppe, K. (2019) Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. *Trends Plant Sci.* 24, 152–164
95. Shi, Y. *et al.* (2016) Unmanned aerial vehicles for high-throughput phenotyping and agronomic research. *PLoS ONE* 11, e0159781
96. Yang, G. *et al.* (2017) Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. *Front. Plant Sci.* 8, 1111
97. Dobbels, A.A. and Lorenz, A.J. (2019) Soybean iron deficiency chlorosis high throughput phenotyping using an unmanned aircraft system. *Plant Methods* 15, 97
98. Huang, H. *et al.* (2018) A fully convolutional network for weed mapping of unmanned aerial vehicle (UAV) imagery. *PLoS ONE* 13, e0196302
99. Altas, Z. *et al.* (2018) Determination of sugar beet leaf spot disease level (*Cercospora Beticola Sacc.*) with image processing technique by using drone. *Curr. Investig. Agric. Curr. Res.* 5, 621–631
100. Bellvert, J. *et al.* (2014) Mapping crop water stress index in a 'Pinot-noir' vineyard: comparing ground measurements with thermal remote sensing imagery from an unmanned aerial vehicle. *Precis. Agric.* 15, 361–376
101. Su, J. *et al.* (2018) Wheat yellow rust monitoring by learning from multispectral UAV aerial imagery. *Comput. Electron. Agric.* 155, 157–166
102. Dang, L.M. *et al.* (2018) UAV based wilt detection system via convolutional neural networks. *Sustain. Comput.Inform. Syst.* Published online May 22, 2018. https://doi.org/10.1016/j.suscom.2018.05.010
103. Wilkinson, M.D. *et al.* (2016) The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3, 1600018
104. Cobb, J. *et al.* (2013) Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Int. J. Plant Breed. Res.* 126, 867–887
105. Nagasubramanian, K. *et al.* (2018) Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. *Plant Methods* 14, 86

106. Kuska, M. *et al.* (2015) Hyperspectral phenotyping on the microscopic scale: towards automated characterization of plant–pathogen interactions. *Plant Methods* 11, 28
107. Pan, S.J. and Yang, Q. (2010) A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359
108. Kolesnikov, A. *et al.* (2019) Big Transfer (BiT): general visual representation learning. *ArXiv.* Published online December 24, 2019. https://arxiv.org/abs/1912.11370
109. Chen, T. *et al.* (2020) Big self-supervised models are strong semi-supervised learners. *ArXiv.* Published online June 17, 2020. https://arxiv.org/abs/2006.10029
110. Zhai, X. *et al.* (2019) A large-scale study of representation learning with the visual task adaptation benchmark. *ArXiv.* Published online October 1 2019. https://arxiv.org/abs/1910.04867
111. Barbedo, J.G.A. (2018) Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Comput. Electron. Agric.* 153, 46–53
112. Cruz, A.C. *et al.* (2017) X-FIDO: an effective application for detecting olive quick decline syndrome with deep learning and data fusion. *Front. Plant Sci.* 8
113. Liu, B. *et al.* (2017) Identification of apple leaf diseases based on deep convolutional neural networks. *Symmetry* 10, 11
114. Kamilaris, A. and Prenafeta-Boldú, F.X. (2018) Deep learning in agriculture: a survey. *Comput. Electron. Agric.* 147, 70–90
115. Settles, B. (2011) From theories to queries: active learning in practice. *Proc. Mach. Learn. Res.* 16, 1–18
116. Awada, L. *et al.* (2018) The adoption of automated phenotyping by plant breeders. *Euphytica* 214, 1–15
117. Han, X. *et al.* (2018) Measurement and calibration of plant-height from fixed-wing UAV images. *Sensors* 18, 4092
118. Gómez-Candón, D. *et al.* (2014) Assessing the accuracy of mosaics from unmanned aerial vehicle (UAV) imagery for precision agriculture purposes in wheat. *Precis. Agric.* 15, 44–56
119. Thomasson, J.A. *et al.* (2019) UAV data reliability improvements based on multifunctional GCPs. In *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping IV*, pp. 110080A, International Society for Optics and Photonics
120. Yu, K. *et al.* (2017) An image analysis pipeline for automated classification of imaging light conditions and for quantification of wheat canopy cover time series in field phenotyping. *Plant Methods* 13, 15
121. Jeong, Y. *et al.* (2018) Cost-effective reflectance calibration method for small UAV images. *Int. J. Remote Sens.* 39, 7225–7250
122. Torres-Sánchez, J. *et al.* (2018) Assessing UAV-collected image overlap influence on computation time and digital surface model accuracy in olive orchards. *Precis. Agric.* 19, 115–133
123. Roitsch, T. *et al.* (2019) Review: new sensors and data-driven approaches – a path to next generation phenomics. *Plant Sci.* 282, 2–10
124. Giuffrida, M.V. *et al.* (2018) Citizen crowds and experts: observer variability in image-based plant phenotyping. *Plant Methods* 14, 12
125. Wiesner-Hanks, T. *et al.* (2018) Image set for deep learning: field images of maize annotated with disease symptoms. *BMC Res. Notes* 11, 440
126. Selvaraj, M.G. *et al.* (2019) AI-powered banana diseases and pest detection. *Plant Methods* 15, 92
127. Barbedo, J.G.A. *et al.* (2018) Annotated plant pathology databases for image-based detection and recognition of diseases. *IEEE Latin Am. Trans.* 16, 1749–1757
128. Barbedo, J.G.A. (2016) A review on the main challenges in automatic plant disease identification based on visible range images. *Biosyst. Eng.* 144, 52–60
129. Arsenovic, M. *et al.* (2019) Solving current limitations of deep learning based approaches for plant disease detection. *Symmetry* 11, 939
130. Arad, B. *et al.* (2020) NTIRE 2020 challenge on spectral reconstruction from an RGB image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 446–447, IEEE
131. Shoeiby, M. et al. (2018) PIRM2018 challenge on spectral image super-resolution: methods and results. In *Computer Vision – ECCV 2018 Workshops* (Leal-Taixé, L. and Roth, S., eds), Springer International Publishing
132. Wang, M. and Deng, W. (2018) Deep visual domain adaptation: a survey. *Neurocomputing* 312, 135–153
133. Everitt, B. (2006) *The Cambridge Dictionary of Statistics* (3rd edn), Cambridge University Press